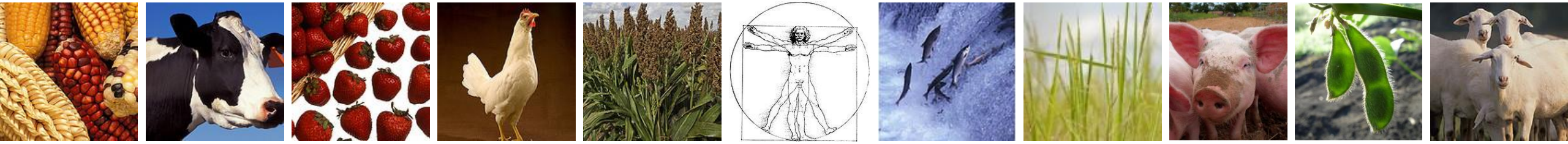


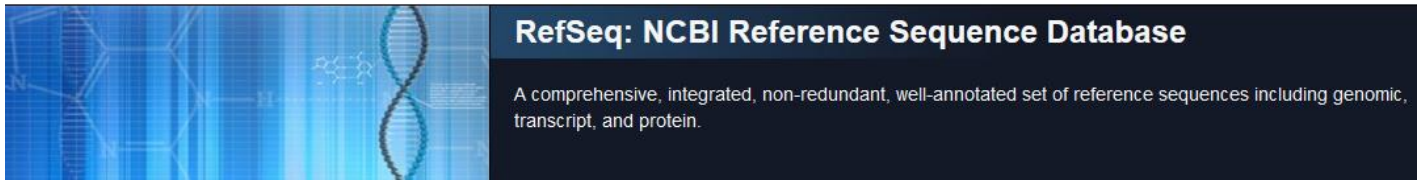
# Navigating NCBI Resources for Plant Genomics

Anjana R Vatsan(raina@nih.gov)

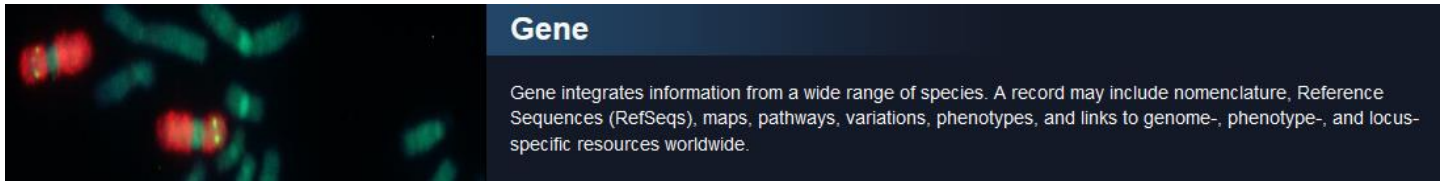
Functional Genomics Workshop

PAG XXVII January 13, 2019

U.S. National Library of Medicine
*National Center for Biotechnology Information*

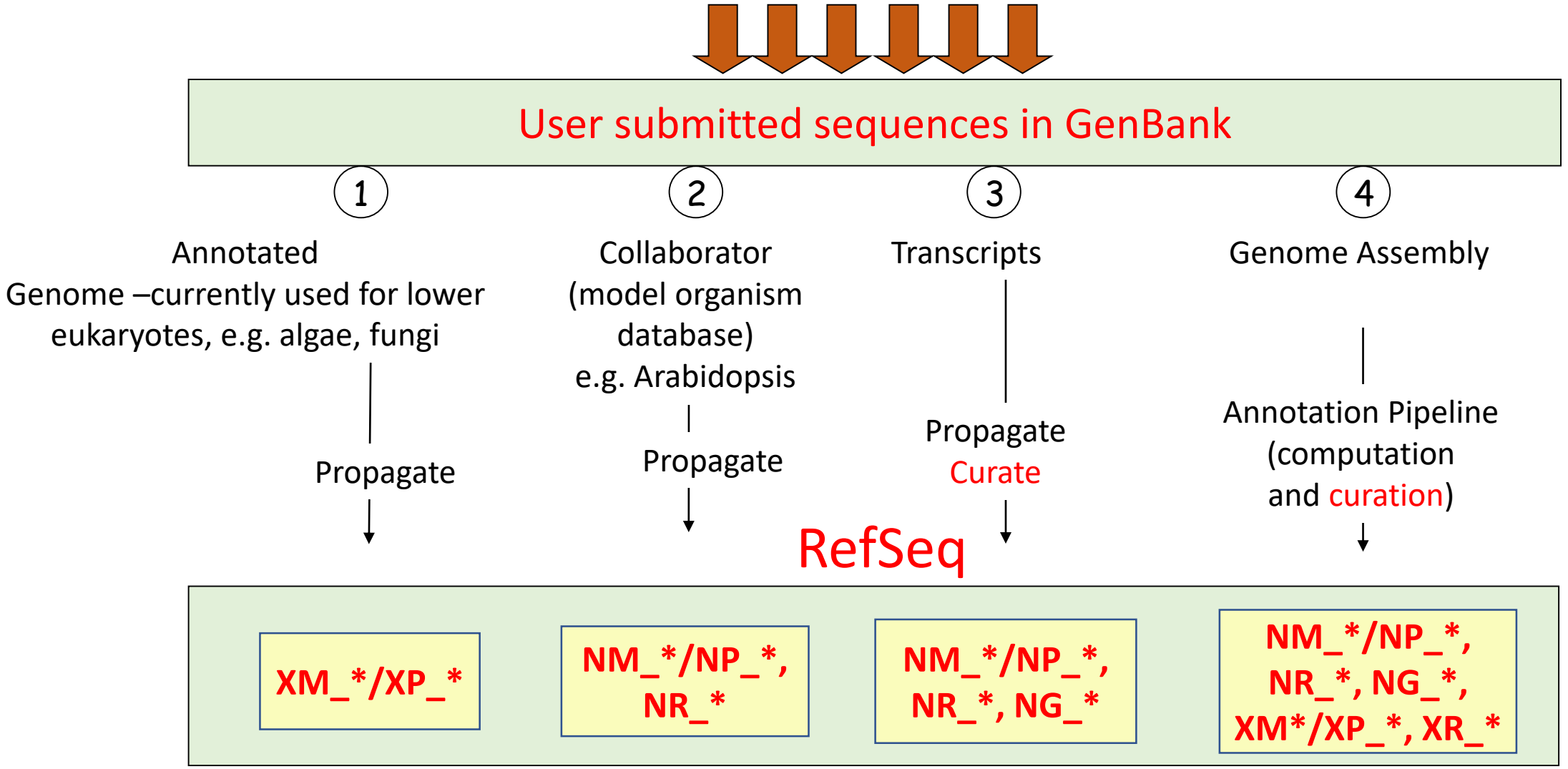



**Reference sequence database (RefSeq)** represents the genomic, transcript and protein products of a gene




**Gene database** contains gene records associated with reference sequences and integrates data from various internal and external resources.

# How is RefSeq created

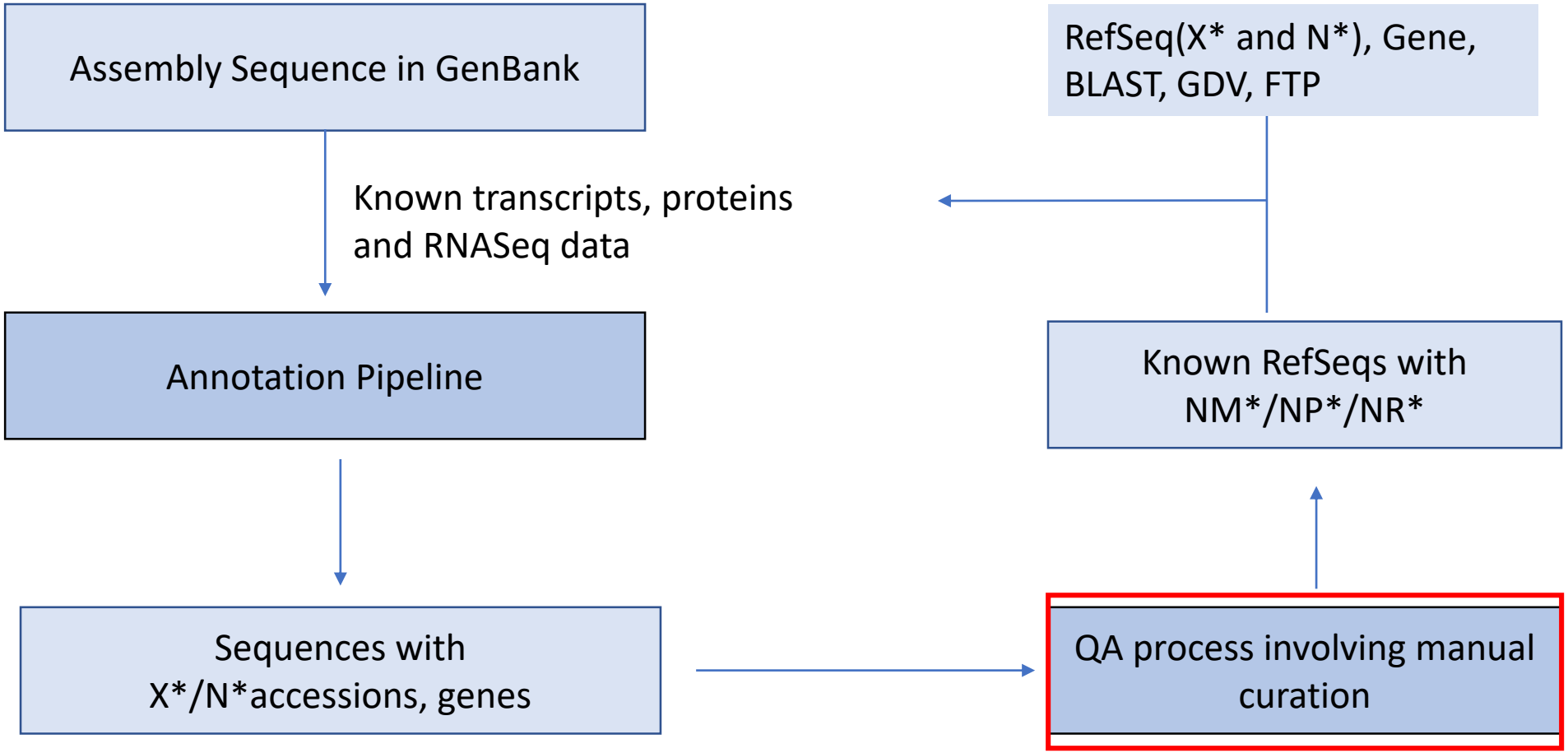


# What is the difference between N* and X* accessions

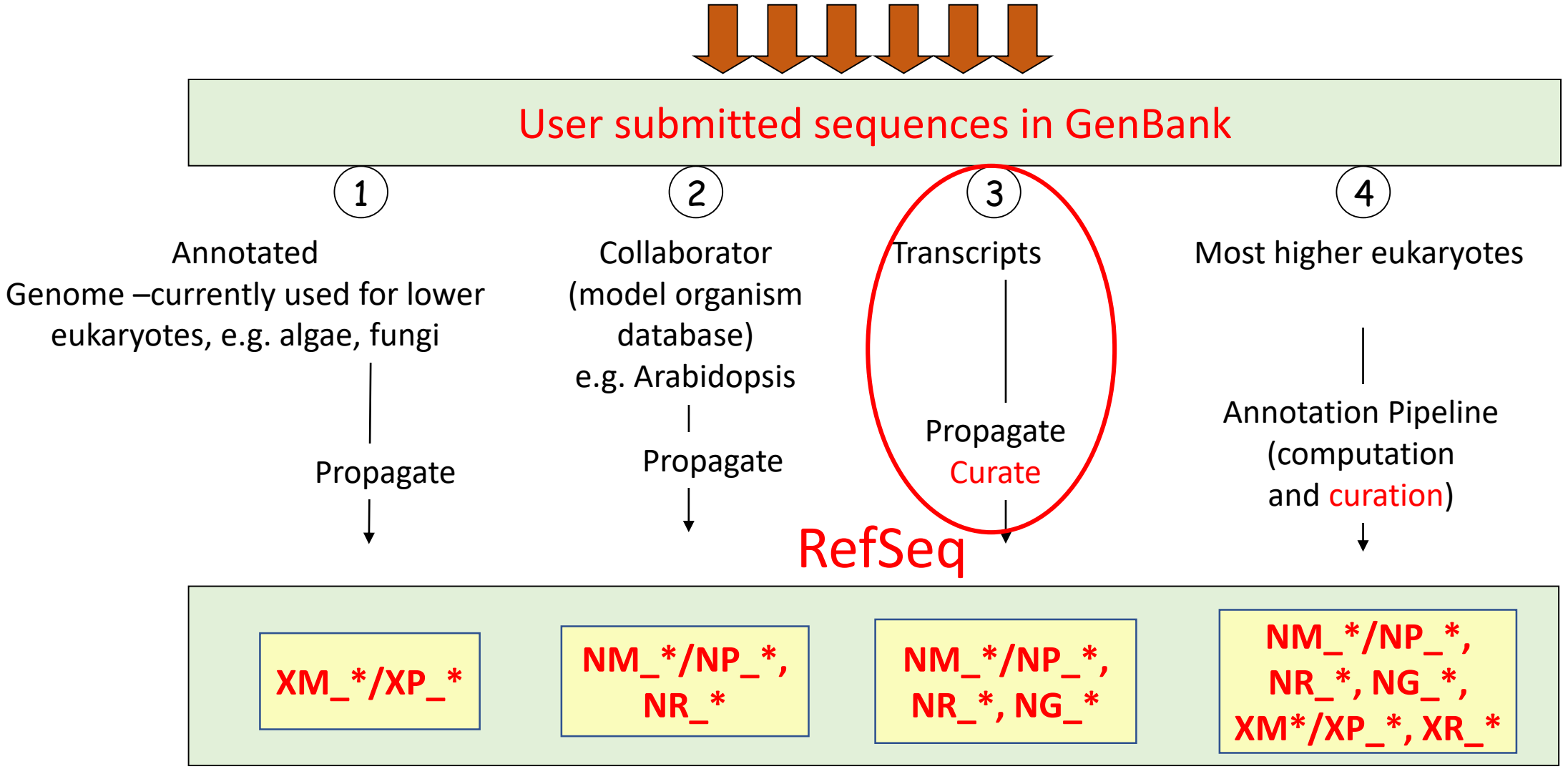
- N* accessions are used for known genes which were traditionally created based on known transcripts.

- X* accessions are predicted models. The vast majority of XMs are fully supported by experimental evidence, and for most species they are on par, quality-wise, with the NMs.

# Annotation process

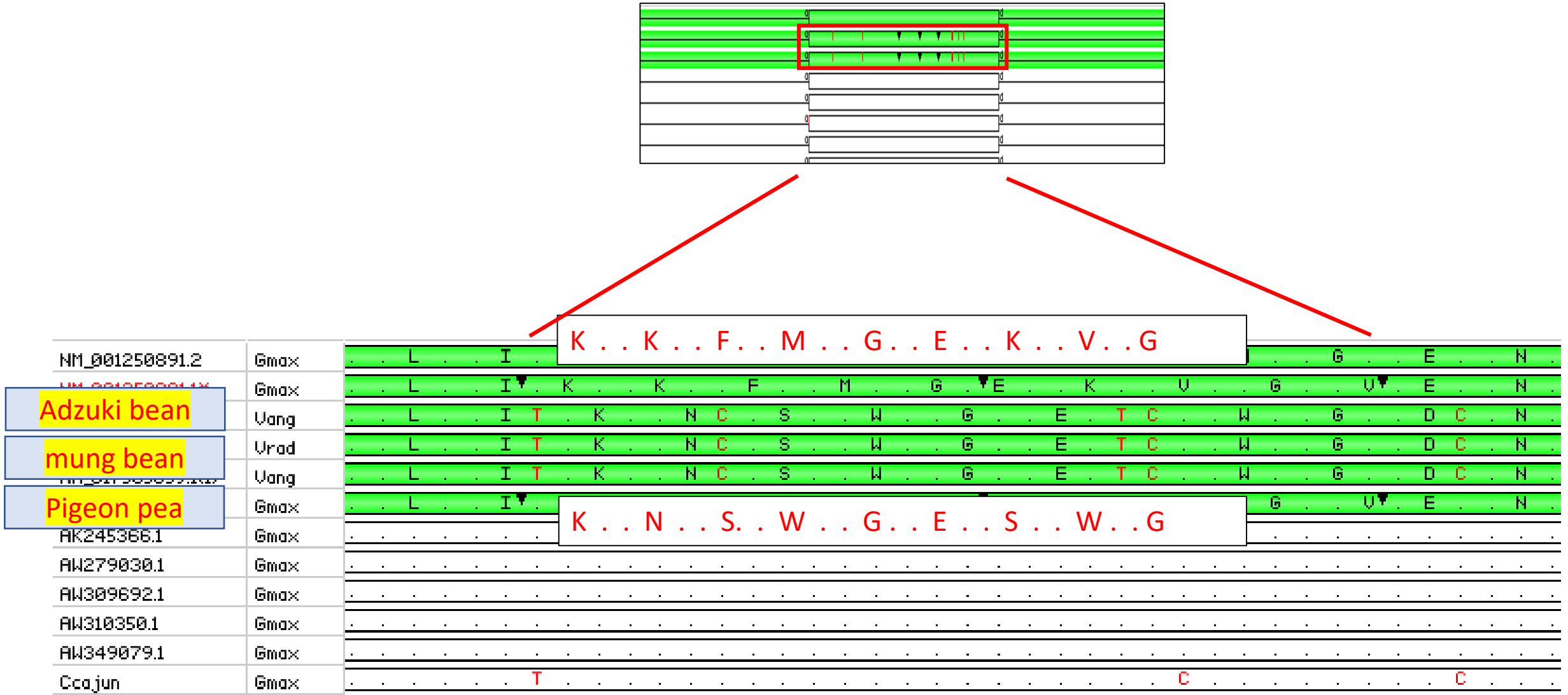


# Annotation Process Flow



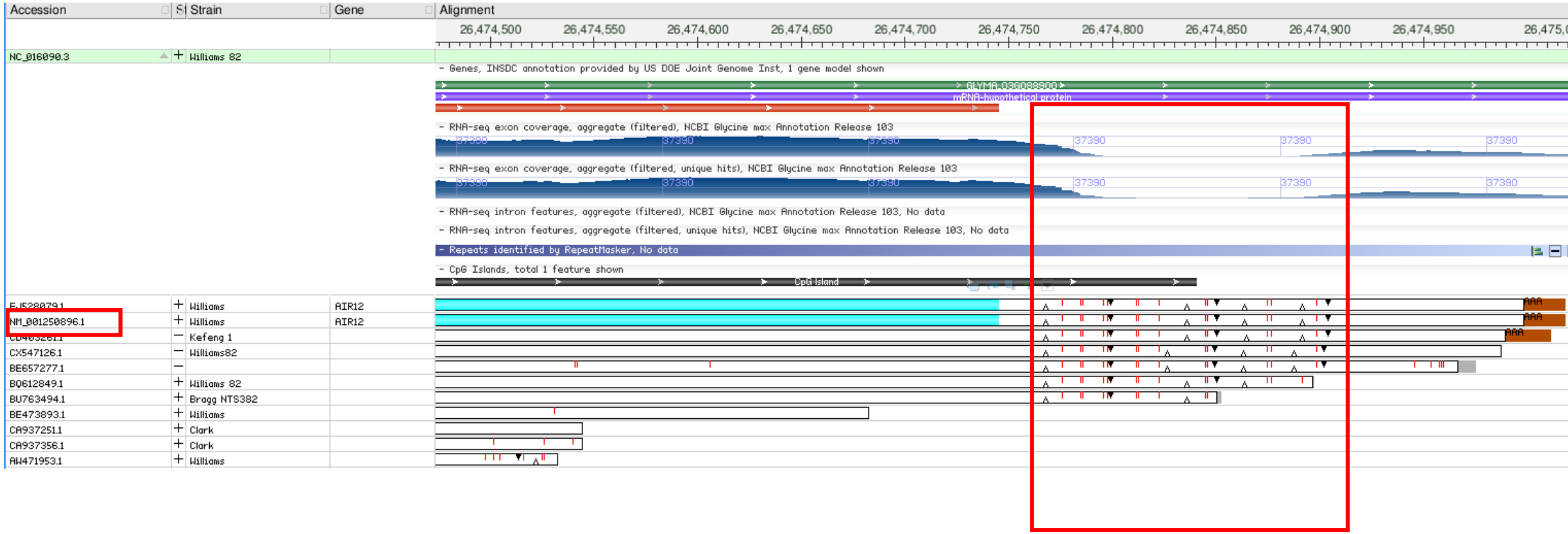
# RefSeq curation

## Correcting Indels/mismatches



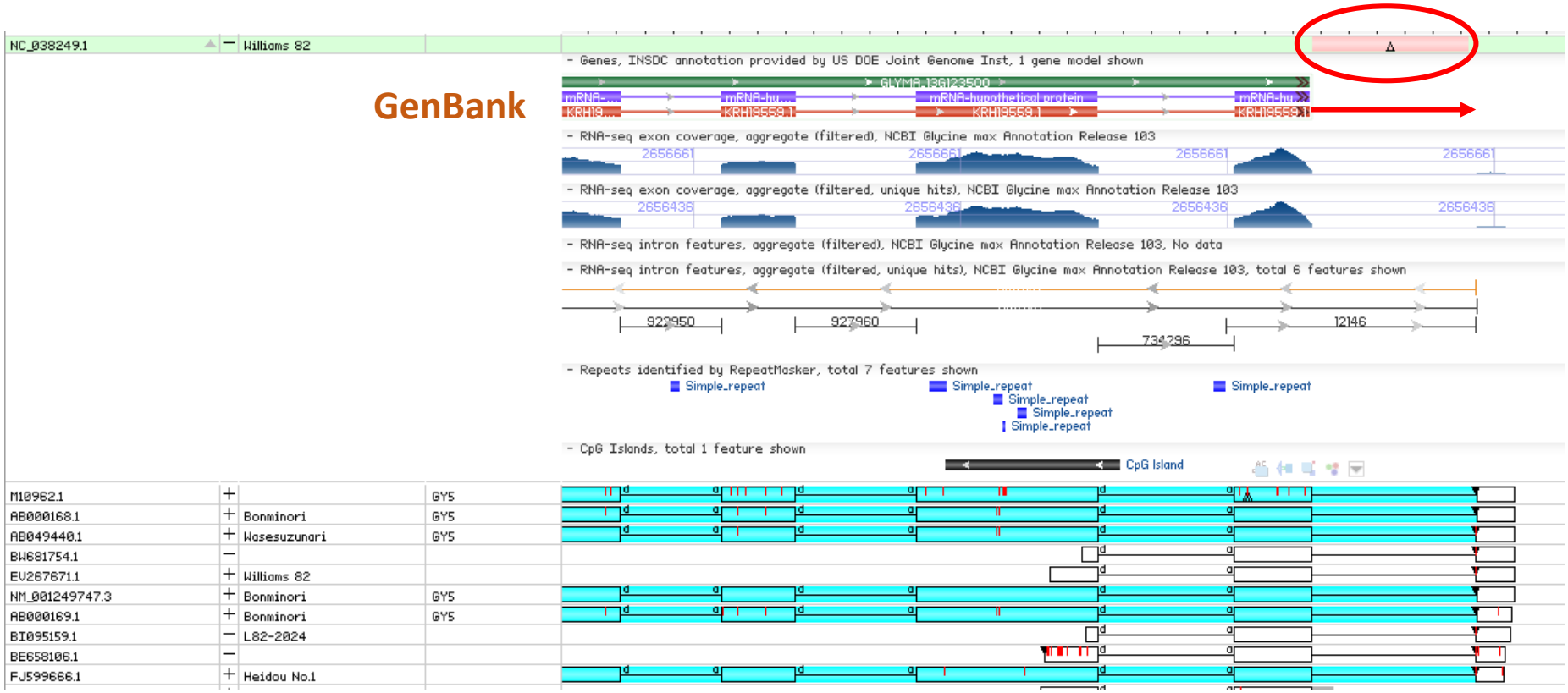
# RefSeq Curation

Transcript based Gene models prevent propagating assembly errors

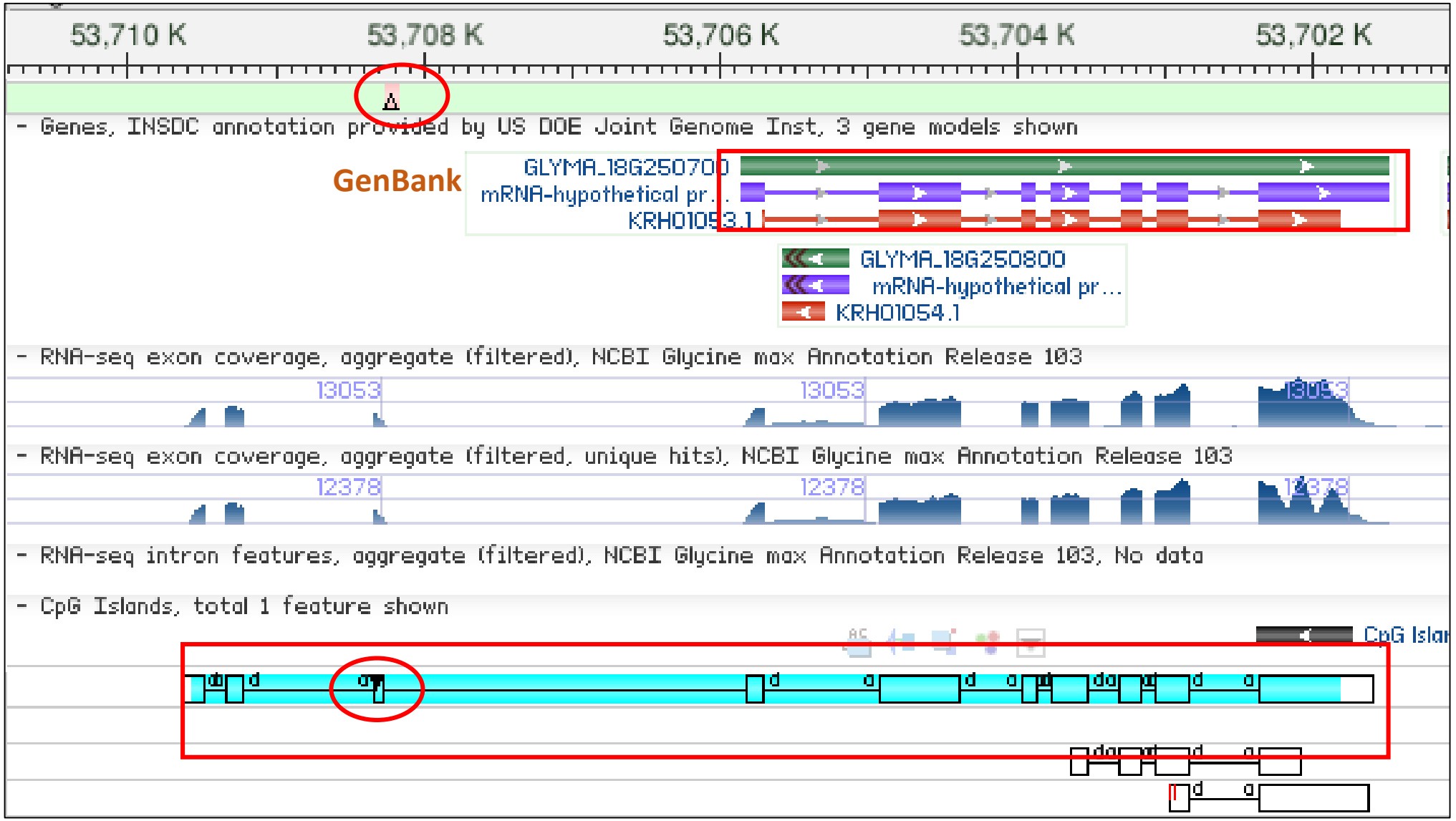


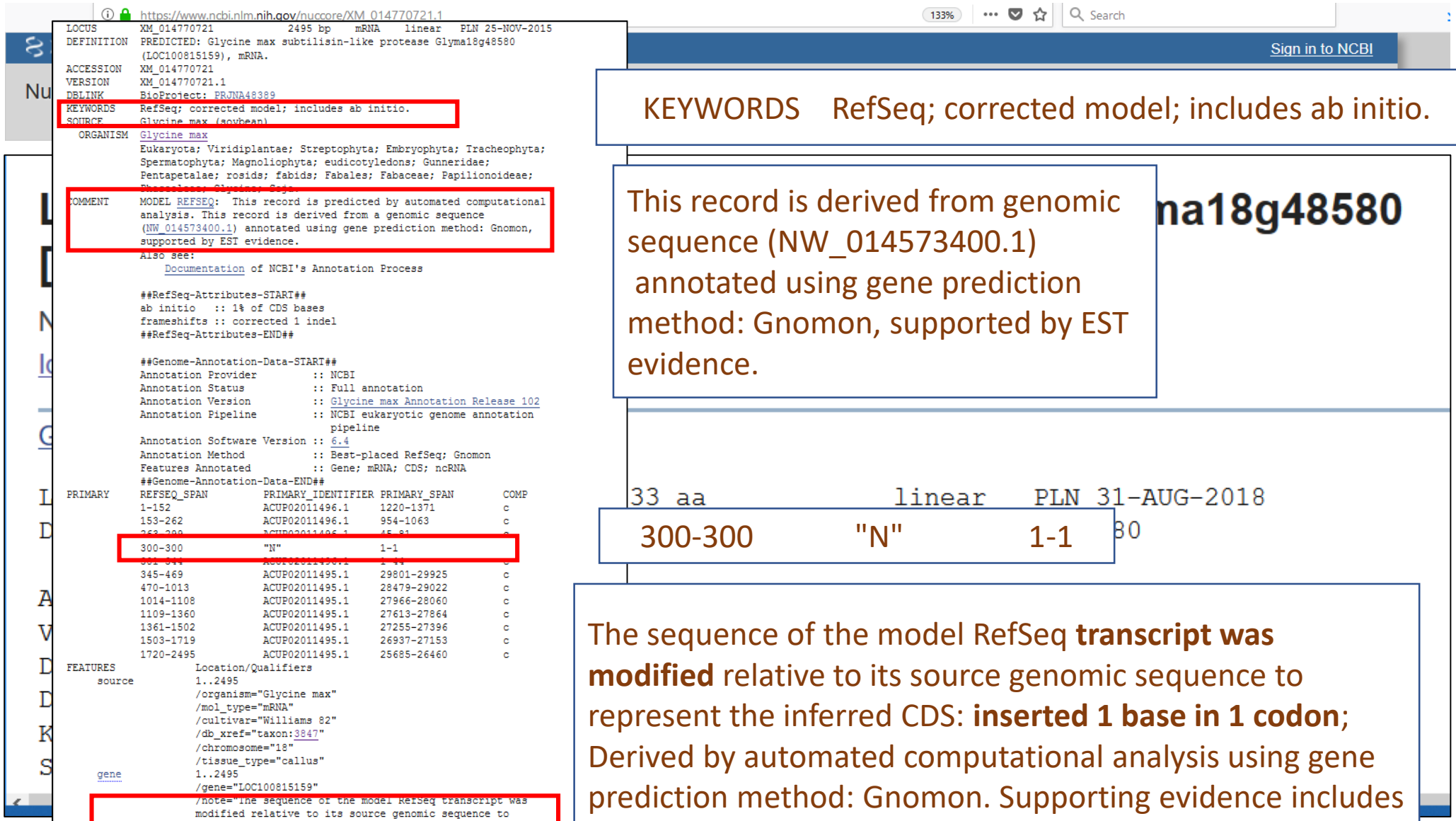
# Corrected Gene Models

## Annotation across gap to give a full representation



# Corrected Gene Models





```
LOCUS       XM_014770721            2495 bp    mRNA    linear   PLN 25-NOV-2015
DEFINITION  PREDICTED: Glycine max subtilisin-like protease Glyma18g48580
            (LOC100815159), mRNA.
ACCESSION   XM_014770721
VERSION     XM_014770721.1
DBLINK      BioProject: PRJNA48389
KEYWORDS    RefSeq; corrected model; includes ab initio.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; fabids; Fabales; Fabaceae; Papilionoideae;
            Phaseoleae; Glycine; Soja.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NW_014573400.1) annotated using gene prediction method: Gnomon,
            supported by EST evidence.
            Also see:
                Documentation of NCBI's Annotation Process

            ##RefSeq-Attributes-START##
            ab initio    :: 1% of CDS bases
            frameshifts :: corrected 1 indel
            ##RefSeq-Attributes-END##

            ##Genome-Annotation-Data-START##
            Annotation Provider         :: NCBI
            Annotation Status           :: Full annotation
            Annotation Version          :: Glycine max Annotation Release 102
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 6.4
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
PRIMARY     REFSEQ_SPAN         PRIMARY_IDENTIFIER PRIMARY_SPAN       COMP
            1-152               ACUP02011496.1     1220-1371          c
            153-262             ACUP02011496.1     954-1063           c
            263-299             ACUP02011496.1     45-81              c
            300-300             "N"                1-1
            301-344             ACUP02011496.1     1-44               c
            345-469             ACUP02011495.1     29801-29925        c
            470-1013            ACUP02011495.1     28479-29022        c
            1014-1108           ACUP02011495.1     27966-28060        c
            1109-1360           ACUP02011495.1     27613-27864        c
            1361-1502           ACUP02011495.1     27255-27396        c
            1503-1719           ACUP02011495.1     26937-27153        c
            1720-2495           ACUP02011495.1     25685-26460        c
FEATURES             Location/Qualifiers
     source          1..2495
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /chromosome="18"
                     /tissue_type="callus"
     gene            1..2495
                     /gene="LOC100815159"
                     /note="The sequence of the model RefSeq transcript was
                     modified relative to its source genomic sequence to
                     represent the inferred CDS: inserted 1 base in 1 codon;
                     Derived by automated computational analysis using gene
                     prediction method: Gnomon. Supporting evidence includes
                     similarity to: 9 ESTs, 8 Proteins, and 96% coverage of the
                     annotated genomic feature by RNAseq alignments
                     /db_xref="GeneID:100815159"
     CDS             63..2264
                     /gene="LOC100815159"
                     /note="The sequence of the model RefSeq protein was
                     modified relative to its source genomic sequence to
```

KEYWORDS RefSeq; corrected model; includes ab initio.

This record is derived from genomic sequence (NW_014573400.1) annotated using gene prediction method: Gnomon, supported by EST evidence.

na18g48580

33 aa linear PLN 31-AUG-2018

300-300 "N" 1-1

The sequence of the model RefSeq **transcript was modified** relative to its source genomic sequence to represent the inferred CDS: **inserted 1 base in 1 codon**; Derived by automated computational analysis using gene prediction method: Gnomon. Supporting evidence includes similarity to: **9 ESTs, 8 Proteins, and 96% coverage of the annotated genomic feature by RNAseq alignments.**

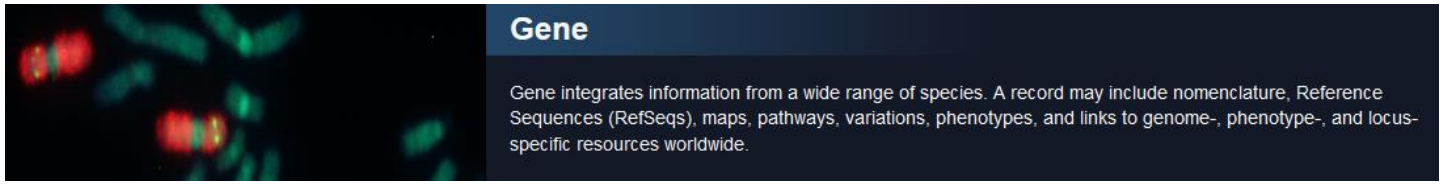
/product="LOW QUALITY PROTEIN: subtilisin-like protease.

# RefSeq curation

Based on PMID: 25192697*, HPPD gene (GeneID:100101901) has two transcription start sites, leading to two HPPD polypeptides, short isoform and long isoform

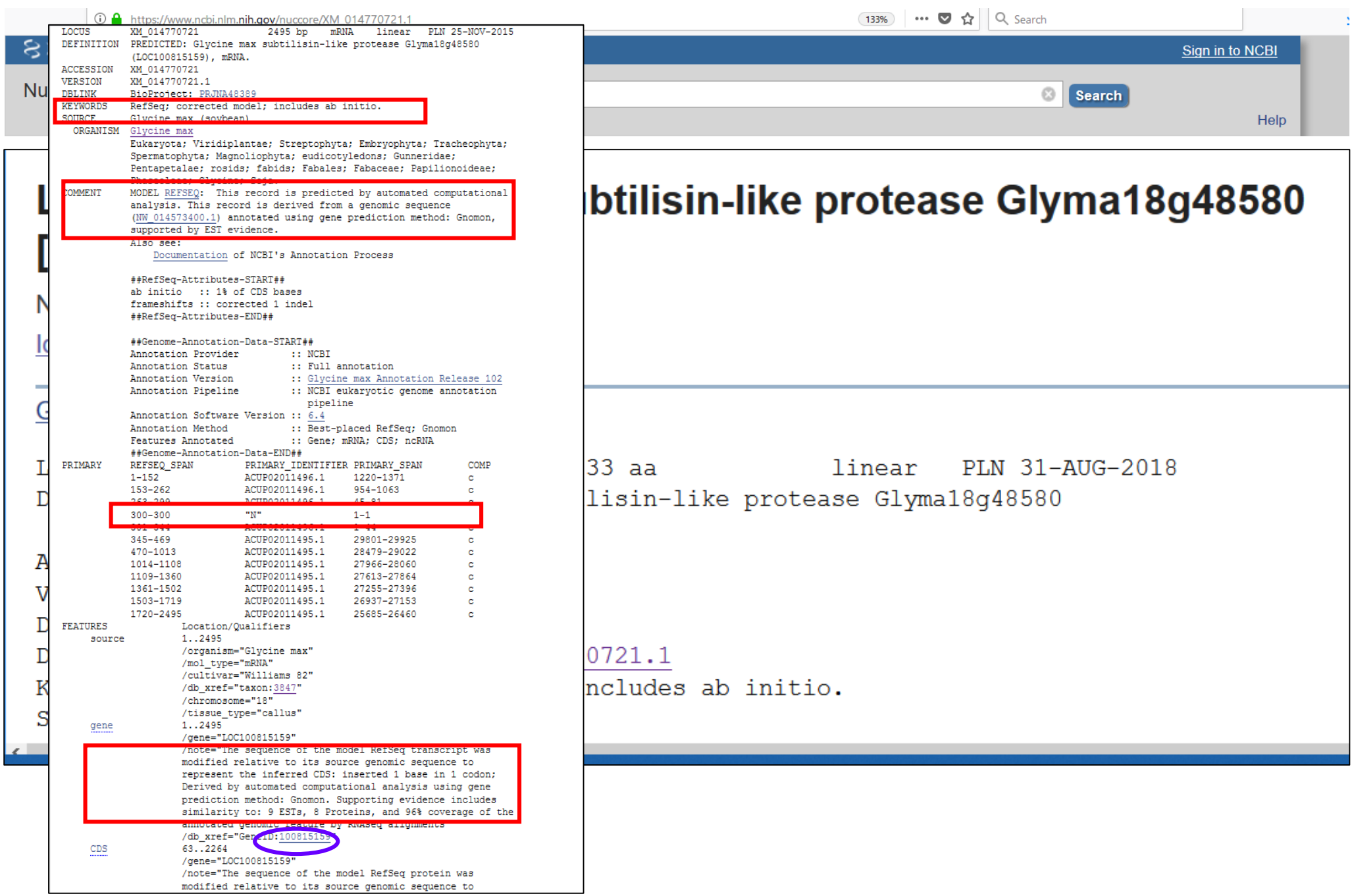

*** Seihl et. al. Plant Physiol. 2014 Nov;166(3):1162-1176. Broad 4-hydroxyphenylpyruvate dioxygenase inhibitor herbicide tolerance in soybean with an optimized enzyme and expression cassette.**

- Update Gene boundary, add/extend UTRs
- Update locus-type, whether coding, non-coding, or pseudogene
- Update product names
- Add publications
- Respond to user feedback




**Gene database** contains gene records associated with reference sequences and integrates data from various internal and external resources.

From the RefSeq record



https://www.ncbi.nlm.nih.gov/nuccore/XM_014770721.1

133%

Search

Sign in to NCBI

Search

Help

```
LOCUS       XM_014770721            2495 bp    mRNA    linear   PLN 25-NOV-2015
DEFINITION  PREDICTED: Glycine max subtilisin-like protease Glyma18g48580
            (LOC100815159), mRNA.
ACCESSION   XM_014770721
VERSION     XM_014770721.1
DBLINK      BioProject: PRJNA48389
KEYWORDS    RefSeq; corrected model; includes ab initio.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; fabids; Fabales; Fabaceae; Papilionoideae;
            Phaseoleae; Glycine; Soja.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NW_014573400.1) annotated using gene prediction method: Gnomon,
            supported by EST evidence.
            Also see:
                Documentation of NCBI's Annotation Process

            ##RefSeq-Attributes-START##
            ab initio    :: 1% of CDS bases
            frameshifts :: corrected 1 indel
            ##RefSeq-Attributes-END##

            ##Genome-Annotation-Data-START##
            Annotation Provider         :: NCBI
            Annotation Status           :: Full annotation
            Annotation Version          :: Glycine max Annotation Release 102
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 6.4
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
PRIMARY     REFSEQ_SPAN         PRIMARY_IDENTIFIER PRIMARY_SPAN       COMP
            1-152               ACUP02011496.1     1220-1371          c
            153-262             ACUP02011496.1     954-1063           c
            263-299             ACUP02011496.1     45-81              c
            300-300             "N"                1-1
            301-344             ACUP02011496.1     1-44               c
            345-469             ACUP02011495.1     29801-29925        c
            470-1013            ACUP02011495.1     28479-29022        c
            1014-1108           ACUP02011495.1     27966-28060        c
            1109-1360           ACUP02011495.1     27613-27864        c
            1361-1502           ACUP02011495.1     27255-27396        c
            1503-1719           ACUP02011495.1     26937-27153        c
            1720-2495           ACUP02011495.1     25685-26460        c
FEATURES             Location/Qualifiers
     source          1..2495
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /chromosome="18"
                     /tissue_type="callus"
     gene            1..2495
                     /gene="LOC100815159"
                     /note="The sequence of the model RefSeq transcript was
                     modified relative to its source genomic sequence to
                     represent the inferred CDS: inserted 1 base in 1 codon;
                     Derived by automated computational analysis using gene
                     prediction method: Gnomon. Supporting evidence includes
                     similarity to: 9 ESTs, 8 Proteins, and 96% coverage of the
                     annotated genomic feature by RNAseq alignments
                     /db_xref="GeneID:100815159"
     CDS             63..2264
                     /gene="LOC100815159"
                     /note="The sequence of the model RefSeq protein was
                     modified relative to its source genomic sequence to
```

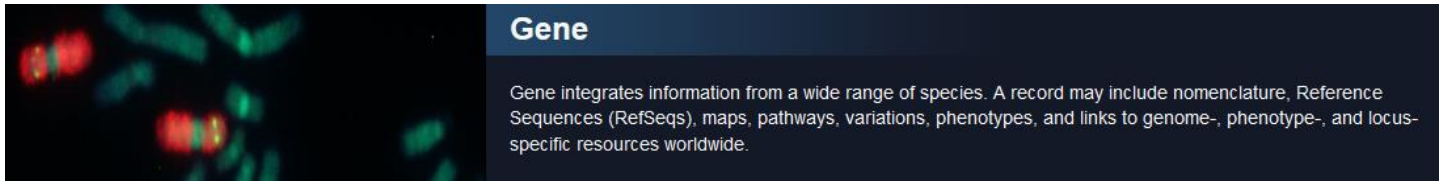
btilisin-like protease Glyma18g48580

33 aa  linear  PLN 31-AUG-2018

lisin-like protease Glyma18g48580

0721.1

ncludes ab initio.



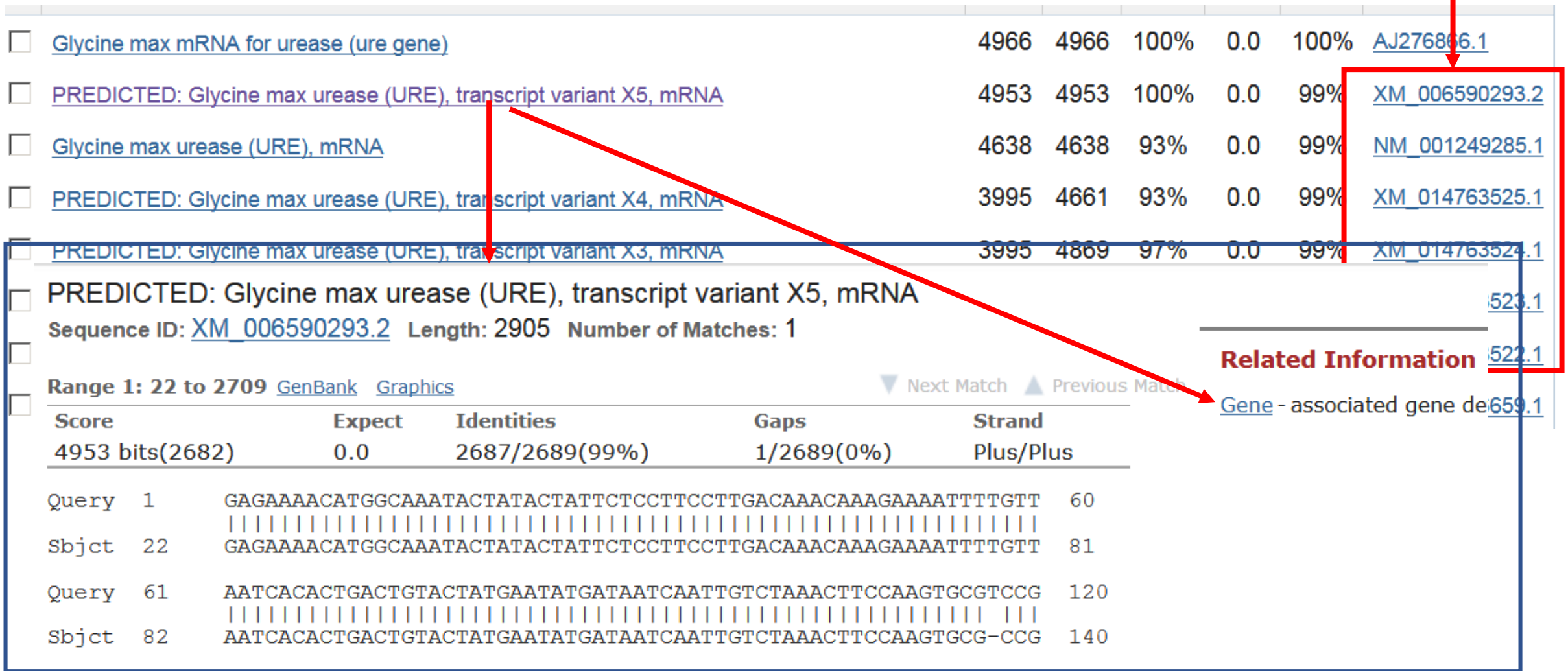

**Gene database** contains gene records associated with reference sequences and integrates data from various internal and external resources.

From the RefSeq record

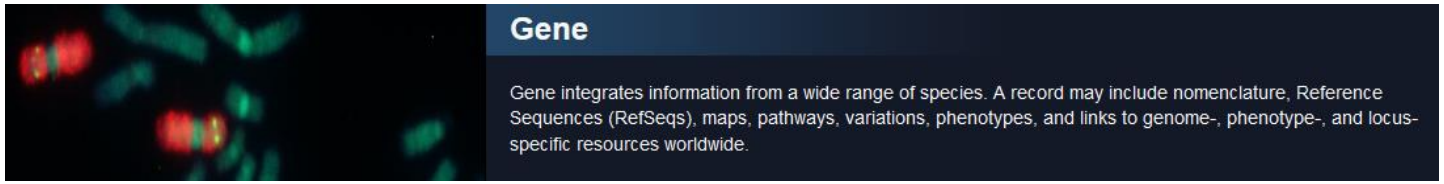
BLAST

Accessing Gene from BLAST output

Select RefSeqs



| | | | | | | |
|---|---|---|---|---|---|---|
| Glycine max mRNA for urease (ure gene) | 4966 | 4966 | 100% | 0.0 | 100% | AJ276866.1 |
| PREDICTED: Glycine max urease (URE), transcript variant X5, mRNA | 4953 | 4953 | 100% | 0.0 | 99% | XM_006590293.2 |
| Glycine max urease (URE), mRNA | 4638 | 4638 | 93% | 0.0 | 99% | NM_001249285.1 |
| PREDICTED: Glycine max urease (URE), transcript variant X4, mRNA | 3995 | 4661 | 93% | 0.0 | 99% | XM_014763525.1 |
| PREDICTED: Glycine max urease (URE), transcript variant X3, mRNA | 3995 | 4869 | 97% | 0.0 | 99% | XM_014763524.1 |




**Gene database** contains gene records associated with reference sequences and integrates data from various internal and external resources.

From the RefSeq record

BLAST

Entrez Gene

Gene

Gene | soybean[orgn] AND Pollen-specific protein SF3 | Search

Create RSS · Create alert · Advanced

Help

# LOC100809121 Pollen-specific protein SF3 [Glycine max (soybean)]

Gene ID: 100809121, updated on 28-Dec-2018

## Summary

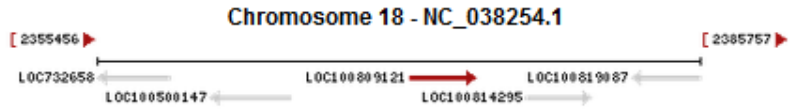
## Genomic context

**Location:** chromosome: 18

See LOC100809121 in Genome Data Viewer

**Exon count:** 7

| Annotation release | Status | Assembly | Chr | Location |
|---|---|---|---|---|
| 103 | current | Glycine_max_v2.1 (GCF_000004515.5) | 18 | NC_038254.1 (2371167..2374487) |
| 102 | previous assembly | Glycine_max_v2.0 (GCF_000004515.4) | 18 | NC_016105.2 (2371163..2374487) |

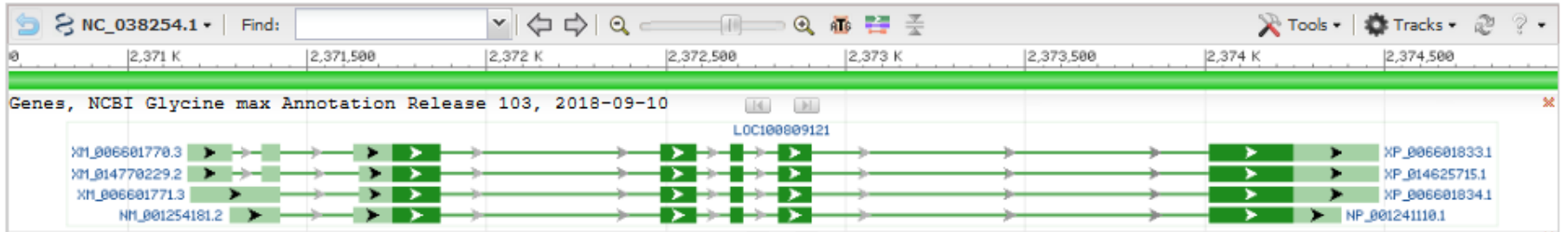



## Genomic regions, transcripts, and products

Go to reference sequence details

**Genomic Sequence:** NC_038254.1 Chromosome 18 Reference Glycine_max_v2.1 Primary Assembly

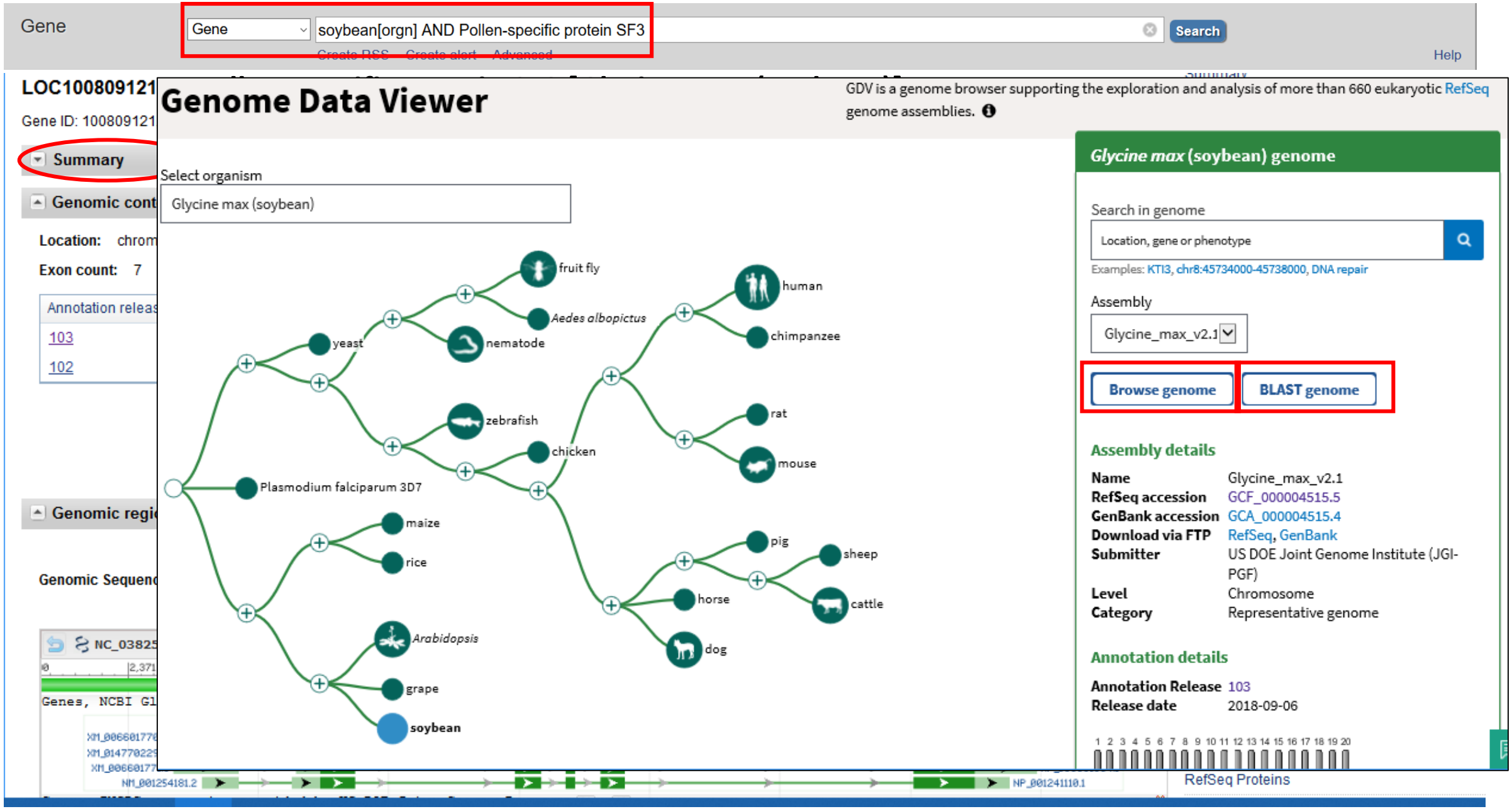
**Go to nucleotide:** Graphics FASTA GenBank

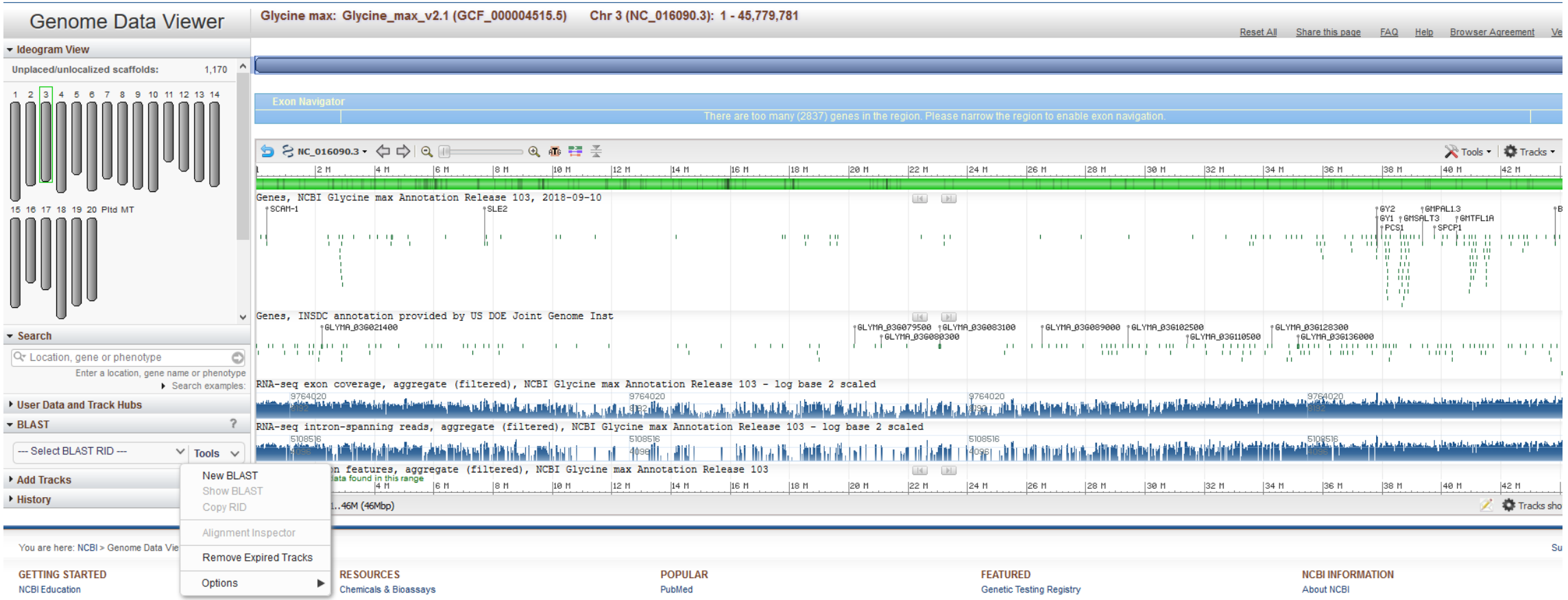



Summary
Genomic context
Genomic regions, transcripts, and products
Bibliography
Variation
General gene information
Homology
General protein information
NCBI Reference Sequences (RefSeq)
Related sequences
Additional links

**Genome Browsers**
Genome Data Viewer

**Related information**
BioProjects
Conserved Domains
EST
Functional Class
Gene neighbors
Genome
Nucleotide
Protein
RefSeq Proteins


Gene
Gene
soybean[orgn] AND Pollen-specific protein SF3
Search
Create RSS
Create alert
Advanced
Help
LOC100809121
Gene ID: 100809121
Summary
Genomic cont
Location: chrom
Exon count: 7
Annotation releas
103
102
Genomic regi
Genomic Sequenc
Genome Data Viewer
GDV is a genome browser supporting the exploration and analysis of more than 660 eukaryotic RefSeq genome assemblies.
Select organism
Glycine max (soybean)
fruit fly
Aedes albopictus
nematode
yeast
zebrafish
chicken
human
chimpanzee
rat
mouse
Plasmodium falciparum 3D7
maize
rice
Arabidopsis
grape
soybean
pig
sheep
horse
cattle
dog
Glycine max (soybean) genome
Search in genome
Location, gene or phenotype
Examples: KTI3, chr8:45734000-45738000, DNA repair
Assembly
Glycine_max_v2.1
Browse genome
BLAST genome
Assembly details
Name Glycine_max_v2.1
RefSeq accession GCF_000004515.5
GenBank accession GCA_000004515.4
Download via FTP RefSeq, GenBank
Submitter US DOE Joint Genome Institute (JGI-PGF)
Level Chromosome
Category Representative genome
Annotation details
Annotation Release 103
Release date 2018-09-06
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
RefSeq Proteins


Genome Data Viewer
Glycine max: Glycine_max_v2.1 (GCF_000004515.5) Chr 3 (NC_016090.3): 1 - 45,779,781
Reset All
Share this page
FAQ
Help
Browser Agreement
Ideogram View
Unplaced/unlocalized scaffolds: 1,170
1 2 3 4 5 6 7 8 9 10 11 12 13 14
15 16 17 18 19 20 Pltd MT
Exon Navigator
There are too many (2837) genes in the region. Please narrow the region to enable exon navigation.
NC_016090.3
Tools
Tracks
Genes, NCBI Glycine max Annotation Release 103, 2018-09-10
SCAM-1
SLE2
GY2
GMPAL13
GY1
GMSALT3
GMTFL1A
PCS1
SPCP1
Genes, INSDC annotation provided by US DOE Joint Genome Inst
GLYMA_03G021400
GLYMA_03G079500
GLYMA_03G083100
GLYMA_03G080300
GLYMA_03G089000
GLYMA_03G102500
GLYMA_03G110500
GLYMA_03G128300
GLYMA_03G136000
RNA-seq exon coverage, aggregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq intron-spanning reads, aggregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
features, aggregate (filtered), NCBI Glycine max Annotation Release 103
Search
Location, gene or phenotype
Enter a location, gene name or phenotype
Search examples:
User Data and Track Hubs
BLAST
--- Select BLAST RID ---
Tools
New BLAST
Show BLAST
Copy RID
Alignment Inspector
Remove Expired Tracks
Options
Add Tracks
History
Tracks sho
You are here: NCBI > Genome Data Vie
GETTING STARTED
NCBI Education
RESOURCES
Chemicals & Bioassays
POPULAR
PubMed
FEATURED
Genetic Testing Registry
NCBI INFORMATION
About NCBI

Gene

Gene | soybean[orgn] AND Pollen-specific protein SF3 | Search

Create RSS  Create alert  Advanced  Help

# Annotation Report

report provides:

- Annotation Release information: The name of the release, important dates, the software version
- Assemblies: A brief description of the annotated assembly(ies)
- Gene and feature statistics: The counts and characteristics of the annotated features
- Alignment of the annotated proteins to a set of high-quality proteins: The number of annotated proteins with hits to quality proteins
- Masking of genomic sequence: How much of the genome was masked
- Transcript and protein alignments: The number and type of evidence retrieved from public databases and used fo
- Similarity of current and previous assembly: The similarity of the current and previous assembly
- Comparison of the current and previous annotations: What proportion of the genes changed in this annotation

more information on the annotation process, please visit the NCBI Eukaryotic Genome Annotation Pipeline page.

notation Release information


RefSeq Proteins

Gene

Gene | soybean[orgn] AND Pollen-specific protein SF3 | Search

Create RSS | Create alert | Advanced | Help

# LOC100809121 Pollen-specific protein SF3 [Glycine max (soybean)]

Gene ID: 100809121, updated on 28-Dec-2018

## Summary

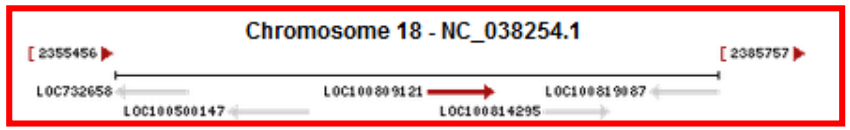
## Genomic context

**Location:** chromosome: 18

See LOC100809121 in Genome Data Viewer

**Exon count:** 7

| Annotation release | Status | Assembly | Chr | Location |
|---|---|---|---|---|
| 103 | current | Glycine_max_v2.1 (GCF_000004515.5) | 18 | NC_038254.1 (2371167..2374487) |
| 102 | previous assembly | Glycine_max_v2.0 (GCF_000004515.4) | 18 | NC_016105.2 (2371163..2374487) |

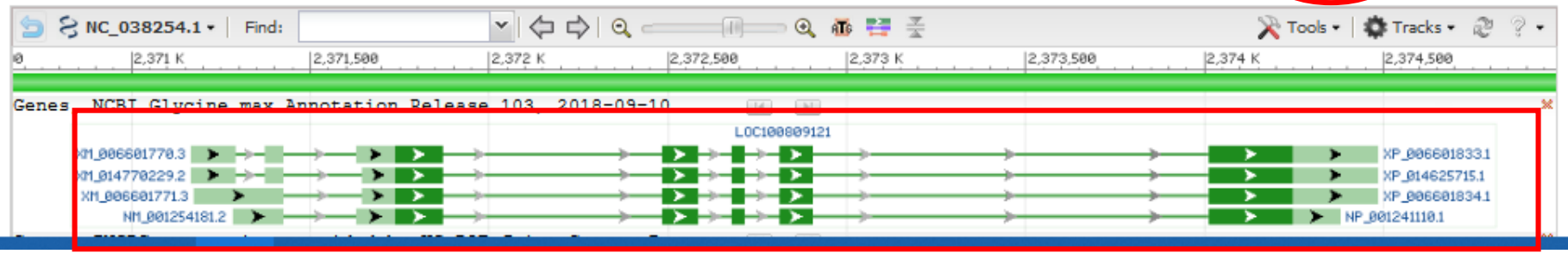



## Genomic regions, transcripts, and products

Go to reference sequence details

**Genomic Sequence:** NC_038254.1 Chromosome 18 Reference Glycine_max_v2.1 Primary Assembly

**Go to nucleotide:** Graphics | FASTA | GenBank




Summary
Genomic context
Genomic regions, transcripts, and products
Bibliography
Variation
General gene information
Homology
General protein information
NCBI Reference Sequences (RefSeq)
Related sequences
Additional links

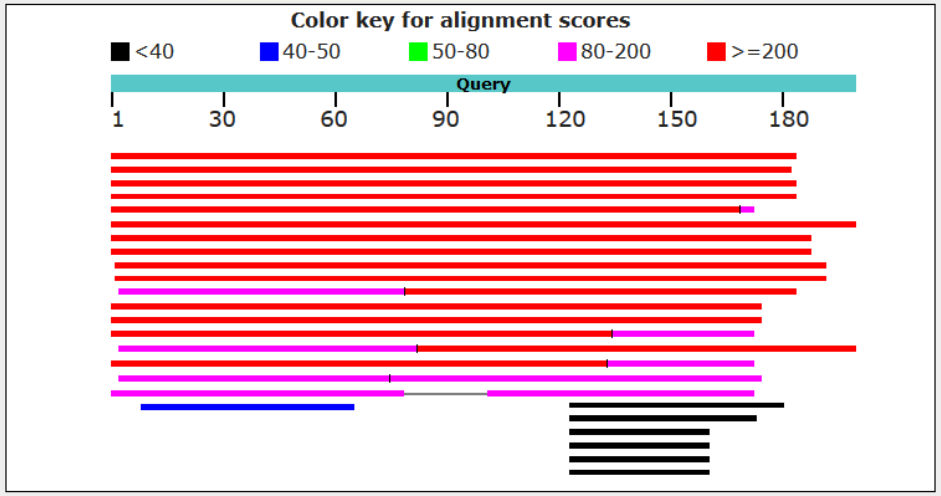
**Genome Browsers**
Genome Data Viewer

**Related information**
BioProjects
Conserved Domains
EST
Functional Class
Gene neighbors
Genome
Nucleotide
Protein
RefSeq Proteins


XM_006595813.2
XM_014766921.1
XM_014766922.1
XP_006595876.1
XP_014622407.1
XP_014622408.1
Genes, INSDC annotation provided by US DOE Joint Genome Inst
GLYMA_14G062700 ...
RNA-seq intron features, aggregate (filtered), NCBI Glycine max An...
Warning: No track data found in this range
RNA-seq exon coverage, Cotyledon, 4 dap, rep 2 (Glycine max, SAMN03083678, filtered), NCBI Glycine max Annotation Release 103 - log base 2 sc...
cotyledon
RNA-seq exon coverage, Flower1 (Glycine max, missing, SAMN02644812, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
flower
RNA-seq exon coverage, developing seed (Glycine max, SAMN04563582, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
seed
RNA-seq exon coverage, fully expanded primary leaves, distal leaf abscission zone (Glycine max, 21 days, SAMN03169132, filtered), NCBI Glycin...
distal leaf
RNA-seq exon coverage, mature leaves in full sunlight, replicate 3 (Glycine max, SAMN02215338, filtered), NCBI Glycine max Annotation Release...
mature leaf
RNA-seq exon coverage, pod, control, block 3 (Glycine max, SAMN02597099, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scal...
pod
RNA-seq exon coverage, seed - globular stage (Glycine max, SAMN02746789, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scal...
seed-globular stage
RNA-seq exon coverage, seeds after 13 days flowering, seeds (Glycine max, SAMN06287523, filtered), NCBI Glycine max Annotation Release 103 - ...
seed
RNA-seq exon coverage, shoot apical meristem after 4-short-day (Glycine max, SAMN02009292, filtered), NCBI Glycine max Annotation Release 103...
shoot apical meristem
Outlier threshold: 5
Other settings:
Fixed Scale
Clip Outliers
Remove track(s)
Configure
Load Defaults
Cancel
Tools
Tracks
Configure Tracks
Configure Track Hubs
NCBI Recommended Track Sets
My NCBI Track Collections
FAQ
40 M
49,042,192
5,100 K
Tracks shown: 6/529


Distribution of the top 36 Blast Hits on 27 subject sequences

Mouse over to see the title, click to show alignments




## Descriptions

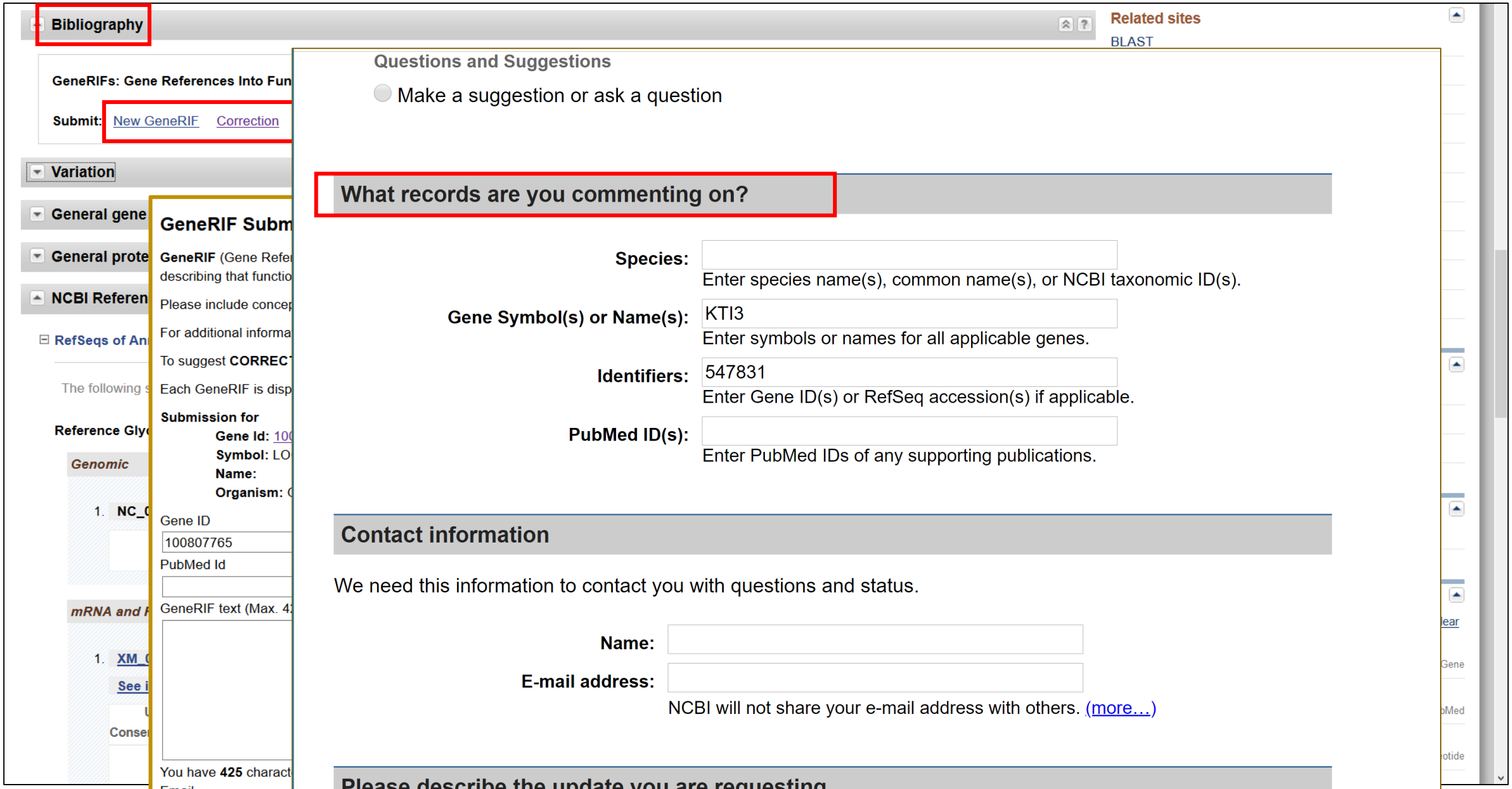
**Sequences producing significant alignments:**

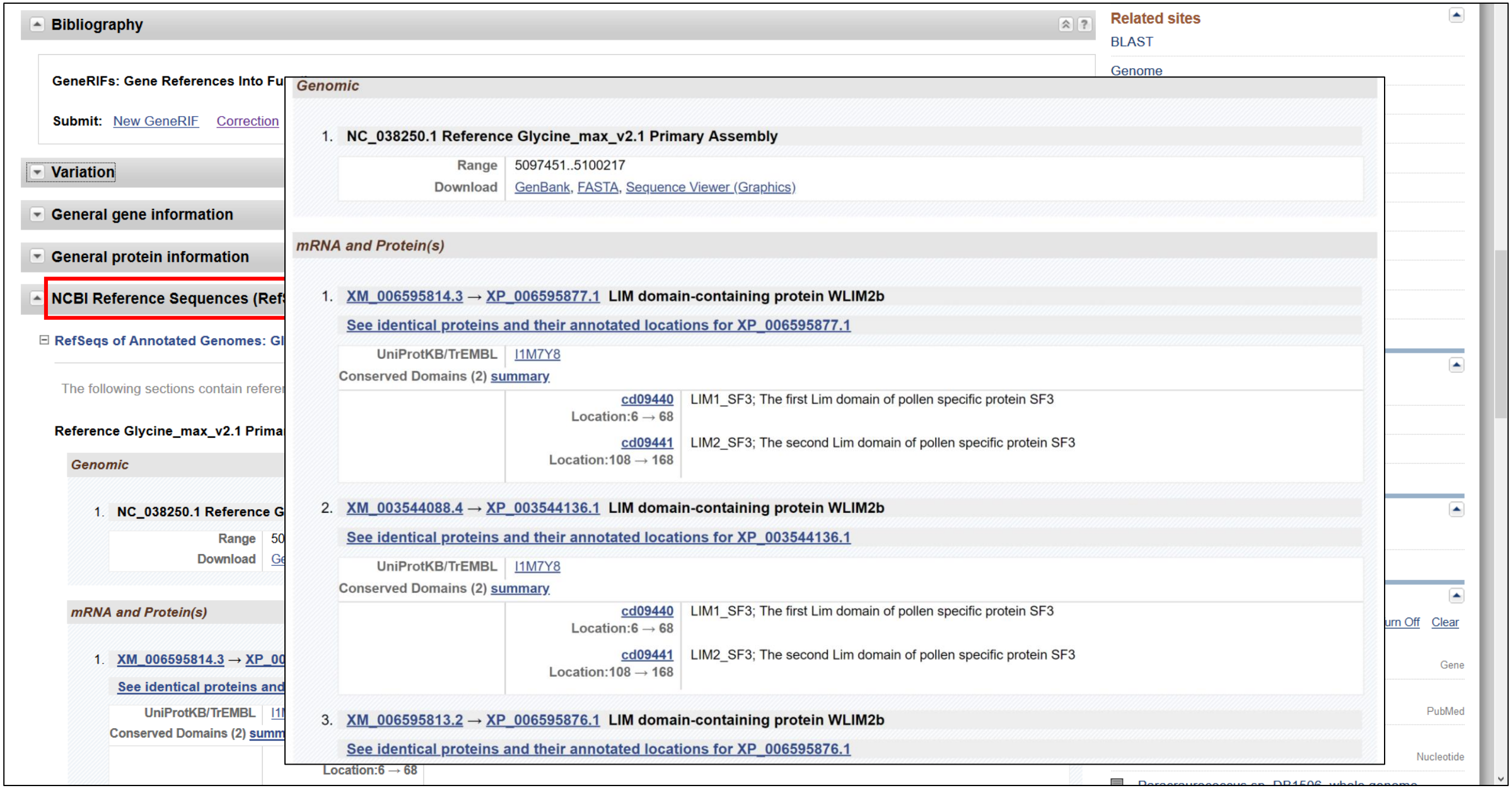
Widely-expressed LIM protein 2B

Select: All None Selected:0

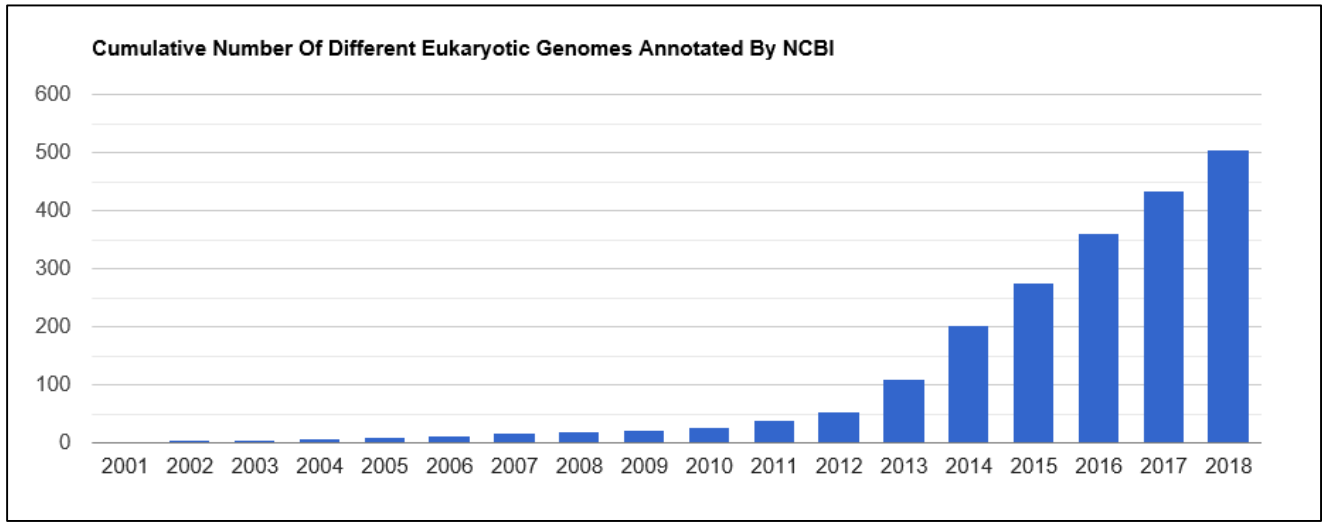
Alignments Download GenPept Graphics Distance tree of results Multiple alignment

| | Description | Max score | Total score | Query cover | E value | Ident | Accession |
|---|---|---|---|---|---|---|---|
| ☐ | GATA type zinc finger transcription factor family protein [Arabidopsis thaliana] | 322 | 322 | 91% | 2e-113 | 82% | NP_191136.1 |
| ☐ | GATA type zinc finger transcription factor family protein [Arabidopsis thaliana] | 322 | 322 | 91% | 4e-113 | 83% | NP_181519.1 |
| ☐ | AT3G55770 [Arabidopsis thaliana] | 319 | 319 | 91% | 3e-112 | 82% | BAH20111.1 |
| ☐ | GATA type zinc finger transcription factor family protein [Arabidopsis thaliana] | 304 | 304 | 91% | 1e-105 | 69% | NP_001190099.1 |

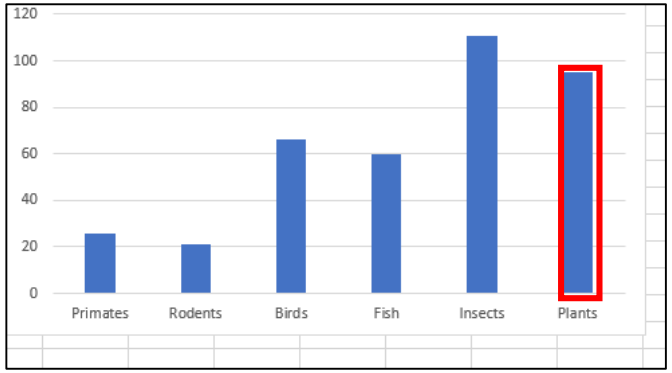

Bibliography
GeneRIFs: Gene References Into Fun
Submit: New GeneRIF Correction
Variation
General gene
General prote
NCBI Referen
RefSeqs of An
The following s
Reference Glyc
Genomic
1. NC_0
mRNA and P
1. XM_0
See i
Conse
Related sites
BLAST
GeneRIF Subm
GeneRIF (Gene Refe
describing that functio
Please include concep
For additional informa
To suggest CORREC
Each GeneRIF is disp
Submission for
Gene Id: 100
Symbol: LO
Name:
Organism:
Gene ID
100807765
PubMed Id
GeneRIF text (Max. 4
You have 425 charact
Email
Submit
Questions and Suggestions
Make a suggestion or ask a question
What records are you commenting on?
Species:
Enter species name(s), common name(s), or NCBI taxonomic ID(s).
Gene Symbol(s) or Name(s):
KTI3
Enter symbols or names for all applicable genes.
Identifiers:
547831
Enter Gene ID(s) or RefSeq accession(s) if applicable.
PubMed ID(s):
Enter PubMed IDs of any supporting publications.
Contact information
We need this information to contact you with questions and status.
Name:
E-mail address:
NCBI will not share your e-mail address with others. (more…)
Please describe the update you are requesting
Notes:


Bibliography
GeneRIFs: Gene References Into Fu
Submit: New GeneRIF Correction
Variation
General gene information
General protein information
NCBI Reference Sequences (Ref
RefSeqs of Annotated Genomes: Gl
The following sections contain referen
Reference Glycine_max_v2.1 Prima
Genomic
1. NC_038250.1 Reference G
Range 50
Download Ge
mRNA and Protein(s)
1. XM_006595814.3 → XP_00
See identical proteins and
UniProtKB/TrEMBL I1
Conserved Domains (2) summ
Location:6 → 68
Related sites
BLAST
Genome
Genomic
1. NC_038250.1 Reference Glycine_max_v2.1 Primary Assembly
Range 5097451..5100217
Download GenBank, FASTA, Sequence Viewer (Graphics)
mRNA and Protein(s)
1. XM_006595814.3 → XP_006595877.1 LIM domain-containing protein WLIM2b
See identical proteins and their annotated locations for XP_006595877.1
UniProtKB/TrEMBL I1M7Y8
Conserved Domains (2) summary
cd09440 LIM1_SF3; The first Lim domain of pollen specific protein SF3
Location:6 → 68
cd09441 LIM2_SF3; The second Lim domain of pollen specific protein SF3
Location:108 → 168
2. XM_003544088.4 → XP_003544136.1 LIM domain-containing protein WLIM2b
See identical proteins and their annotated locations for XP_003544136.1
UniProtKB/TrEMBL I1M7Y8
Conserved Domains (2) summary
cd09440 LIM1_SF3; The first Lim domain of pollen specific protein SF3
Location:6 → 68
cd09441 LIM2_SF3; The second Lim domain of pollen specific protein SF3
Location:108 → 168
3. XM_006595813.2 → XP_006595876.1 LIM domain-containing protein WLIM2b
See identical proteins and their annotated locations for XP_006595876.1
urn Off Clear
Gene
PubMed
Nucleotide

# More than 500 genomes have been annotated Computationally using the NCBI's Annotation Pipeline




95 plants



# Annotation

- Primates (26)
- Rodents (21)
- Even-toed ungulates and whales (Cetartiodactyla) (23)
- Other Mammals (52)
- Birds (66)
- Fish (60)
- Other Vertebrates (19)
- Insects (111)
- Other Invertebrates (33)
- Plants (95)

**FTP** - FTP Download **B** - Organism-specific BLAST **AR** - Annotation Report **GDV** - Genome Data Viewer

| Species | RefSeq assembly(ies) | Annotation Release | Freeze Date | Release Date | Links | | | |
|---|---|---|---|---|---|---|---|---|
| Cicer arietinum (chickpea) | ASM33114v1 (GCF_000331145.1) | 102 | 2018-12-03 | 2018-12-27 | FTP | B | AR | GDV |
| Abrus precatorius (Indian licorice) | Abrus_2018 (GCF_003935025.1) | 100 | 2018-12-18 | 2018-12-20 | FTP | B | AR | |
| Coffea eugenioides (eudicots) | Ceug_1.0 (GCF_003713205.1) | 100 | 2018-12-05 | 2018-12-08 | FTP | B | AR | GDV |
| Coffea arabica (coffee) | Cara_1.0 (GCF_003713225.1) | 100 | 2018-11-26 | 2018-12-04 | FTP | B | AR | GDV |

NCBI Genome Resources Workshop

Location: Town and Country, Pacific Salon 2
Date: Monday, Jan 14 12:50 PM

PO0737 Navigating NCBI Resources for Plant Genomics

NCBI Booth: 223

# Thank you.

RefSeq/Gene

**Terence Murphy**
Eric Cox
Catherine Farrell
Tamara Goldfarb
Diana Haddad
John Jackson
Vinita Joardar
Kelly McGarvey
Michael Murphy
Nuala O'Leary

RefSeq Developers

Alex Astashyn
Olga Ermolaeva
Vamsi Kodali
Craig Wallin

Shashi Pujar
Bhanu Rajput
Sanjida Rangwala
Lillian Riddick
Barbara Robbertse
Brian Smith-White
Pooja Strope
David Webb

Annotation Pipeline

**Francoise Thibaud-Nissen**
Paul Kitts
Mike Dicuccio
Wratko Hlavina
Avi Kimchi

Jinna Choi
Boris Kiryutin
Patrick Masterson
Eyal Mozes
Anton Perkov
Dan Rausch
Robert Smith
Alexandre Souvorov

GDV/Remap/GBench

Valerie Schneider
Peter Meric
Nathan Bouk
Hsiu-Chuan Chen
Cliff Clausen
Anatoliy Kuznetsov

A cast of thousands

Ken Katz
Michael Ovetsky
Lukas Wagner
Andrei Shkeda
Donna Maglott
**Kim Pruitt**
**Jim Ostell**
**David Lipman**

Watch NCBI News for updates!
http://www.ncbi.nlm.nih.gov/news/
https://www.youtube.com/user/NCBINLM

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.